

## APPENDIX C: TECHNICAL NOTES

# DATA USED IN THIS REPORT

## The Survey of Doctorate Recipients

The 1993 Survey of Doctorate Recipients (SDR) includes individuals under 76 years of age who received a research doctorate in science or engineering from a U.S. university in 1992 or earlier. The focus of the current report was restricted to individuals in the labor market<sup>40</sup> at the time of the survey (April 1993). Thus, individuals who were neither employed nor seeking employment at that time were excluded from the analyses. The available sample size was approximately 36,000 cases.

## Historical Data

Changes have been made in the population definition and data collection procedures for the SDR that reduce the direct comparability of the earlier surveys with the 1993 survey.<sup>41</sup> The 1973 data were adjusted to make them as comparable as possible to the 1993 data.<sup>42</sup>

A report by the National Science Foundation (NSF), *Unemployment Rates and Employment Characteristics for Scientists and Engineers, 1971*, is used for comparison purposes within this report, even though the NSF study differed considerably in population definition and research design from the 1993 SDR. The scientists for the earlier NSF survey were those included in the 1970 National Register of Scientific and Technical Personnel. To be included in the register, individuals were required to have “full professional standing based on academic training and work experience, as determined by the appropriate scientific professional society for the fields of science covered.”<sup>43</sup> Approximately 60 percent of the scientists did not have doctorates.<sup>44</sup> Engineers were selected from a mailing list maintained by the Engineers Joint Council that “consisted of 23 major engineering

societies and constituted about 40 percent of the total number of individuals in the Nation identified as engineers.”<sup>45</sup> Thus, the definitions of scientist and engineer in the 1972 study were not strictly comparable to one another, nor were they comparable to the 1993 definition of an individual with a doctorate in one of the science and engineering fields.

## Total Population Data

Information on total population unemployment was taken from data collected by the Bureau of Labor Statistics (BLS) in the Current Population Survey (CPS). The definition of unemployment used in the CPS is essentially the same as that used in the SDR.

## TREND ANALYSIS

Changes in the SDR methodology over time (e.g., fluctuating response rates and population definitions) have affected the size of the unemployment rate estimates. In 1973, the response rate for the survey was 75 percent. Between 1973 and 1989, the response rate gradually declined to 55 percent. In 1991, extensive locating and telephone follow-up procedures were instituted that helped raise the response rate in 1991 and 1993 to approximately 87 percent. During the 1991 redesign of the SDR, the population definition was modified. The 1973 study used a sample frame that included many individuals who received doctoral degrees from non-U.S. institutions. However, after 1973, only individuals with doctoral degrees from U.S. institutions were added to the survey. By 1991, it was clear that the coverage of the non-U.S.-educated population was extremely poor. Since improving this coverage within the SDR was not practical, this segment was deleted entirely.

To understand the likely impact of the 1991 changes on the unemployment rate, rates were calculated for 1989 and 1991 using population and methodological definitions that were as similar as possible. Foreign-educated individuals were excluded from the 1989 estimate, and individuals who responded during

---

<sup>40</sup> An individual in the labor market is defined as employed or, if not employed, having actively sought work within the preceding four months or being on layoff.

<sup>41</sup> See the Technical Notes for a discussion of changes in the SDR over time.

<sup>42</sup> See below (under Trend Analysis) for more information on this adjustment.

<sup>43</sup> NSF 1972, pp. 112–113.

<sup>44</sup> NSF 1972, p. 15.

---

<sup>45</sup> NSF 1972, pp. 114–115.

the telephone follow-up stage in 1991 were also excluded from this comparison. The resulting unemployment rate for the 1989 group was 0.8 percent; the 1991 rate was 1.3 percent. The published rates with the differing population definitions and methodology for the two years were 0.8 percent and 1.4 percent. Therefore, it appears that the changes in methodology and population definition resulted in a slight increase in the estimated unemployment rate between the two years. Since the impact of the changed methodology on unemployment rates did not appear to be substantial, it was decided that a fairly good approximation for trend analysis purposes could be made by adding 0.1 percent to the pre-1991 unemployment rates.

There are some discrepancies in reported doctoral unemployment rates for 1973. The rate reported in the National Academy of Sciences publications was 1.2 percent, although the 1973 rate reported in the NSF's *Characteristics of Doctoral Scientists and Engineers in the United States: 1989* was 1.1 percent. Since the latter rate was published as part of the trend analysis used to calculate the adjusted 1989 rate, it was assumed that the NSF rate was the best rate for use in calculating adjusted pre-1989 figures.

## VARIABLE DEFINITIONS

### Unemployment Rate

The definition of unemployment used in this report is the standard Federal definition of the percent of individuals in the labor force who were not employed. The labor force is defined as individuals who were employed, were on lay-off, or had sought work within the preceding four weeks. Although this is the most commonly used measure of unemployment, other measures are used. The Bureau of Labor Statistics, for example, in a 1995 article discusses a variety of alternative measures used for different purposes (Bregger and Haugen).

### Involuntary Part-Time Rate

The involuntary part-time rate is defined as the number of individuals who reported working part-time exclusively because suitable full-time work was not available, divided by the number of individuals in the labor force.

### Involuntary Out-of-Field Rate

For this report, the involuntary out-of-field rate is defined as the number of individuals (other than those who were involuntarily part-time employed) who reported that they were working out of their doctoral field at least partially because suitable work in the field was not available, divided by the number of individuals in the labor force. This is slightly different than the definition used in the NSF report, *Characteristics of Doctoral Scientists and Engineers in the United States: 1993*, which combines individuals who are involuntarily part-time or involuntarily out-of-field into a single measure, referred to as involuntary out-of-field. For the purposes of this report, the components are broken out. This report also uses the number of individuals in the labor force as the denominator for calculating this rate, rather than the number of employed individuals, in order to facilitate combining the three measures of adverse career events.

### Occupation

Standard SRS occupational groupings were used for coding the 1988 science and engineering occupations. These codes are detailed in NSF 96-302. For non-science and engineering occupations, a further breakdown of occupations into managerial or professional specialty positions was made. Non-management/professional specialty occupations included: technologists and technicians; clerical/administrative support; computer programmers; surveyors; farmers, foresters, and fishermen; nurses; sales and marketing; service occupations other than health; and elementary and secondary teachers. Jobs in this category were selected based on the characteristics of individuals in these jobs in the 1993 National Survey of College Graduates. The remaining non-S&E occupations were considered to be managerial and professional specialty jobs. This category includes the clergy, lawyers, and managers, where high-level degrees are common.

### Variables Related to 1988 Employment and Occupational Status

The 1993 SDR included a series of questions about the employment status of individuals in 1988. These questions asked whether the individual had

changed employer or occupation since 1988 and, if so, asked for information about the 1988 position. This retrospective information was used throughout the report to describe 1988 occupational characteristics.

## Other Variables

In examining associations between single variables and the unemployment rate, the goal was to restrict analyses to groups that consisted of at least 400 sample cases. This is a relatively large cut-off, because of the high sampling variability encountered in small samples when rates are as low as 1.6 percent. Meeting the minimum sample size goal required collapsing categories. When logical combinations did not permit the desired sample size goal to be met, smaller sample sizes were retained. If this was not feasible, small residual categories were treated as missing for the purposes of examining the bivariate relationships between the independent variables of interest and unemployment status.

The categories used in the bivariate analyses were also used as a starting point for creating dummy variables for the multivariate work. However, since the regression routines used in the multivariate analyses ignore all cases with missing values for one or more variables, the missing value codes were examined again before conducting the multivariate analysis. Some categories (for example, "Other Physical Science" under degree field) that were not displayed in the univariate analysis were used in the multivariate analysis. The remaining missing value cases were treated as if they belonged to whichever dummy variable category had been selected for omission in the dummy variable regression. Normally, this was the modal category for the variable.

## STANDARD ERRORS AND TESTS OF SIGNIFICANCE

Observed differences in comparing unstandardized unemployment rates between groups were tested for statistical significance at the .05 significance level. Standard errors for these tests were calculated using the equation appropriate for a simple random sample. This is equivalent to assuming that there is no design effect. Although this methodology provides only an approximate estimate of the standard error, it greatly simplifies

computation. Since the sample design for the survey was a stratified random sample, this approach should provide reasonably good estimates.

Sample sizes for some of the 1973 subgroups used in comparing 1973 and 1993 unemployment rates were not readily available. Therefore, the number of cases in the subgroup was estimated by multiplying the 1993 sample size for the group by the ratio of total 1973 sample size to total 1993 sample size. Although this is a fairly rough test, it provides general guidance on the probable statistical significance of observed differences.

## STANDARDIZATION METHODOLOGY

The first step in developing a model for estimating unemployment was an examination of the bivariate associations between the independent variables of interest and unemployment. Some variables were eliminated from further consideration after examination of these relationships based primarily on whether the observed bivariate relationship could reasonably be interpreted as one in which the independent variable affected unemployment. For example, non-work-related training appeared to be associated with high unemployment rates. However, it seems more reasonable to believe that being unemployed leads one to seek additional training than that obtaining additional training increases the probability of unemployment. The bivariate relationships for these omitted variables were discussed in Appendix B.

The preliminary analysis also suggested the appropriate shape of curves to fit in the multivariate analysis. For example, for variables (such as years since the doctorate was earned) that display high unemployment rates at the extremes of the distribution, parabolic relationships were fit by including squared and linear terms for the relevant independent variables.

Once the preliminary independent variables were identified, a multiple regression analysis was performed to identify possible problems with multicollinearity that required the deletion of additional variables.<sup>46</sup> Stepwise regression analysis was also used to determine if there were additional variables that could be deleted due to a lack of statistical significance. Variables omitted at this

---

<sup>46</sup> All analyses for this report were performed using SAS.

stage included gender, race/ethnicity, and whether the individual had children. At that point, a limited number of plausible two-way interactions were introduced into the analysis and tested for statistical significance (for example, gender by whether children are present). The next step was to perform a logistic regression analysis. The preliminary logistic model was simplified by eliminating variables that were not statistically significant from the model.<sup>47</sup> The parameters for the final logistic regression model are presented in table C-1.

A problem with using logistic regression analysis is that interpretation of the results is not straightforward. The impact of an independent variable on unemployment depends on the value of the other variables in the model. Since such complex relationships are difficult to comprehend, a standardization technique was used. For most variables, iterative techniques were used to select a standardization value for all factors other than the independent variable of interest. This resulted in a total unemployment rate equal to the observed unemployment rate.

The standardization methodology selected was modified slightly to deal with situations where there was a logical dependence between categorical independent variables in the analysis. For example, individuals categorized as not employed in 1988 in the occupational analysis were categorized the same way in the sector variable. Logit regression parameters were calculated for each category formed by cross-classifying the interdependent independent variables. For example, chemists in the private sector would have a combined logit parameter equal to the sum of the parameters for the dummy variable for chemists, the value for the dummy variable for the private sector, and the values of

the dummy variables used to indicate employment and student status. Standardization was performed for these detailed occupations by sector categories, and the value for each sector and occupation was obtained by weighting these subcategories—for example, sector categories within an occupation group—according to the observed distribution.

An exception to this general treatment was made for the variable characterizing 1988 occupation by the percent of individuals within the occupation in the 1993 NSCG who were involuntarily out-of-field (an indicator of the perceived desirability of the occupation). For this variable, the standardized values for the two unemployed in 1988 categories were set equal to the values observed in the analysis of 1988 occupation and 1988 sector described prior to calculating standardized values for the remaining categories.

For continuous variables, standardization was done within categories. For the purpose of evaluating the regression values, the midpoint of the category was used to estimate the dependent variable mean unless knowledge of the data suggested a different value would be more appropriate.

In standardizing for disability status, the categories were not mutually exclusive, because of the possibility that an individual could have multiple disabilities. Instead of standardizing to the total observed unemployment rate or forcing the categories to be mutually exclusive, unemployment rates were standardized to a hypothetical total unemployment rate calculated from the observed values of the univariate disability categories.

---

<sup>47</sup> Note that interaction effects were tested after a decision was made on whether the primary variables should be retained.

**Table C-1. Regression parameters and standard errors for logistic regression model**

Page 1 of 4

Independent Variables	Parameters	Standard Error
INTERCEPT.....	-2.4199	0.5168 *
Place of employment or residence		
Connecticut.....	-0.6984	0.1086 *
Massachusetts.....	-0.6324	0.0651 *
Other New England Region.....	-0.2991	0.1029 *
New York.....	-0.7088	0.0534 *
New Jersey.....	-0.8328	0.0687 *
Pennsylvania.....	-0.4133	0.0615 *
Ohio.....	-0.0872	0.0602
Indiana.....	-0.7502	0.1187 *
Illinois.....	-0.7582	0.0732 *
Michigan.....	-0.5932	0.0833 *
Wisconsin.....	-1.3787	0.1645 *
Minnesota.....	-0.2931	0.0892 *
Missouri.....	-1.0868	0.1344 *
Other West North Central Region.....	-1.7205	0.1664 *
District of Columbia.....	-1.2771	0.112 *
Florida.....	-0.02	0.0684
Maryland.....	-0.3117	0.0616 *
North Carolina.....	-0.2645	0.0779 *
Virginia.....	-1.0293	0.096 *
Other South Atlantic Region.....	-1.359	0.1013 *
East South Central Region.....	-0.7796	0.0771 *
Texas.....	-0.2405	0.0524 *
Other West South Central Region.....	-0.4771	0.0873 *
Mountain Region.....	0.0682	0.0458
{California}**.....	--	--
Washington.....	-0.3089	0.0762 *
Other Pacific.....	-0.8448	0.1054 *
Other.....	-0.8977	0.2368 *
Years since receipt of Ph.D.		
Years since receipt of Ph.D.....	0.0604	0.00492 *
Years since receipt of Ph.D. squared.....	0.000696	0.000114 *
Field of degree:		
{Biological sciences}**.....	--	--
Mathematical and computer sciences.....	-0.2194	0.0757 *
Agricultural sciences.....	0.5105	0.0721 *
Geological and environmental sciences.....	0.4827	0.069 *
Chemistry.....	-0.2337	0.0565 *

See explanatory information, if any, and SOURCE at the end of table.

**Table C-1. Regression parameters and standard errors for logistic regression model**

Page 2 of 4

Independent Variables	Parameters	Standard Error
Field of degree (continued):		
Physics and astronomy.....	0.4015	0.0581 *
Other physical sciences.....	0.00927	0.2615
Economics.....	0.3063	0.075 *
Political science.....	0.4099	0.0746 *
Psychology.....	0.1769	0.0522 *
Sociology/Anthropology.....	-0.1228	0.0716
Other social sciences.....	-0.1328	0.0858
Chemical engineering.....	0.0227	0.0847
Civil engineering.....	-1.1496	0.1672 *
Electrical engineering.....	0.1944	0.0719 *
Mechanical engineering.....	-0.5593	0.1126 *
Other engineering.....	0.0996	0.0634
Age when doctorate received		
Age at Ph.D.....	0.1175	0.00243 *
Part-time experience		
Years part-time experience.....	0.0642	0.00627 *
Years part-time experience squared.....	-0.00375	0.000346 *
Full-time experience		
Years full-time experience.....	-0.0959	0.00434 *
Years full-time experience squared.....	0.000861	0.0001 *
Employed in 1988?*		
Employed in 1988?.....	0.3878	0.0818 *
Employed or non-doctorate in 1988?.....	-1.4275	0.0768 *
Occupation in 1988		
Mathematical and computer scientists.....	-0.1023	0.0752
Postsecondary teachers: math/computer.....	-0.9684	0.1227 *
{Biological scientists}**.....	--	--
Agricultural scientists.....	-1.1581	0.126 *
Postsecondary teachers: life sciences.....	-1.2751	0.0972 *
Chemists.....	0.4554	0.0704 *
Geologists and environmental scientists.....	-0.5614	0.1009 *
Physicists.....	-0.2747	0.0826 *
Other physical sciences.....	0.4223	0.1264 *
Postsecondary teachers: physical sciences.....	-1.5963	0.1239 *
Psychologists.....	-1.526	0.0947 *
Other social scientists.....	-0.6794	0.0983 *
Postsecondary teachers: social sciences.....	-0.958	0.0779 *
Electrical engineers.....	-0.6309	0.1019 *

See explanatory information, if any, and SOURCE at the end of table \*

**Table C-1. Regression parameters and standard errors for logistic regression model**

Page 3 of 4

Independent Variables	Parameters	Standard Error
Occupation in 1988 (continued):		
Other engineers.....	0.0534	0.0653
Postsecondary teachers: engineering.....	-1.475	0.1485 *
Managerial and professional1.....	-0.6172	0.0534 *
Other non-scientists and engineers1.....	-0.7309	0.0699 *
Employment sector in 1988		
Medical school.....	-0.679	0.0793 *
University-affiliated research institute.....	0.0251	0.0606
{Other four-year college/university}.....	--	--
Other educational employer.....	-0.0813	0.1
Private for profit employer.....	0.66	0.0391 *
Self-employed -- incorporated.....	-0.4535	0.1052 *
Self-employed -- not incorporated.....	-0.6924	0.0832 *
Private not for profit employer.....	-0.0198	0.0677
State government.....	-0.5452	0.1285 *
U.S. government -- civilian position.....	-0.4476	0.071 *
Other government.....	0.4491	0.0838 *
Parents' education		
Less than high school.....	0.2103	0.0398 *
{High school}.....	--	--
Some college.....	0.2954	0.0385 *
2-year college.....	-0.2939	0.072 *
4-year college.....	-0.0681	0.0391
Some graduate school.....	-0.1282	0.0681
Master's.....	-0.1779	0.049 *
Doctorate.....	0.1038	0.0424 *
Disability status		
{Not disabled}.....	--	--
Disability related to seeing.....	-0.186	0.0878 *
Disability related to hearing.....	0.4292	0.0606 *
Disability related to walking.....	0.3177	0.1161 *
Disability related to lifting.....	0.2902	0.0965 *
Average percent in 1988 occupation involuntarily out of field****.....	6.7056	0.747 *
Foreign research experience		
Conducted foreign research.....	0.2869	0.0433 *
Would consider conducting foreign research.....	0.3483	0.0304 *
{Neither conducted nor would consider conducting foreign research}**.....	--	--

See explanatory information, if any, and SOURCE at the end of table



**Table C-1. Regression parameters and standard errors for logistic regression model**

Page 4 of 4

Independent Variables	Parameters	Standard Error
Marital status		
Married.....	-4.2888	0.5053 *
married.....	--	--
Interactions between gender and marital status		
Married female.....	0.4624	0.0703 *
Other.....	--	--
Interactions between gender and whether children in home		
Women with children in the home.....	0.8661	0.0575 *
Men with children in the home.....	-0.0781	0.032 *
No children in the home (total).....	--	--
Interactions between gender and race/ethnicity		
Non-Hispanic white women.....	-0.8512	0.0583 *
Non-Hispanic black women.....	-0.3166	0.1897
Hispanic women.....	-1.3319	0.2001 *
Asian women.....	-0.2604	0.0972 *
Native American (total)***.....	-1.8798	0.4761 *
Interactions between marital status and race/ethnicity***		
{Married non-Hispanic whites}.....	--	--
Unmarried non-Hispanic whites.....	-3.5736	0.5051 *
Married non-Hispanic blacks.....	-0.6395	0.15 *
Unmarried non-Hispanic blacks.....	-3.7227	0.5244 *
Married Asians.....	-0.053	0.0463
Unmarried Asians.....	-4.311	0.5135 *
Married Hispanics.....	0.3014	0.0986 *
Unmarried Hispanics.....	-3.4297	0.5281 *

\* Statistically significant at the .05 level.

\*\* {}'s are used to indicate omitted dummy regression variables.

\*\*\* Due to the small number of Native Americans in the sample, a single category of Native Americans was used.

\*\*\*\* Unemployed individuals were given a value equal to the mean of employed individuals on this variable.

<sup>1</sup>See the Technical Notes for an explanation of occupations included in this category.

**SOURCE:** National Science Foundation/SRS, 1993 Survey of Doctorate Recipients.